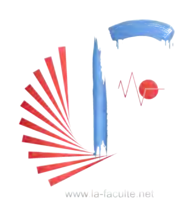
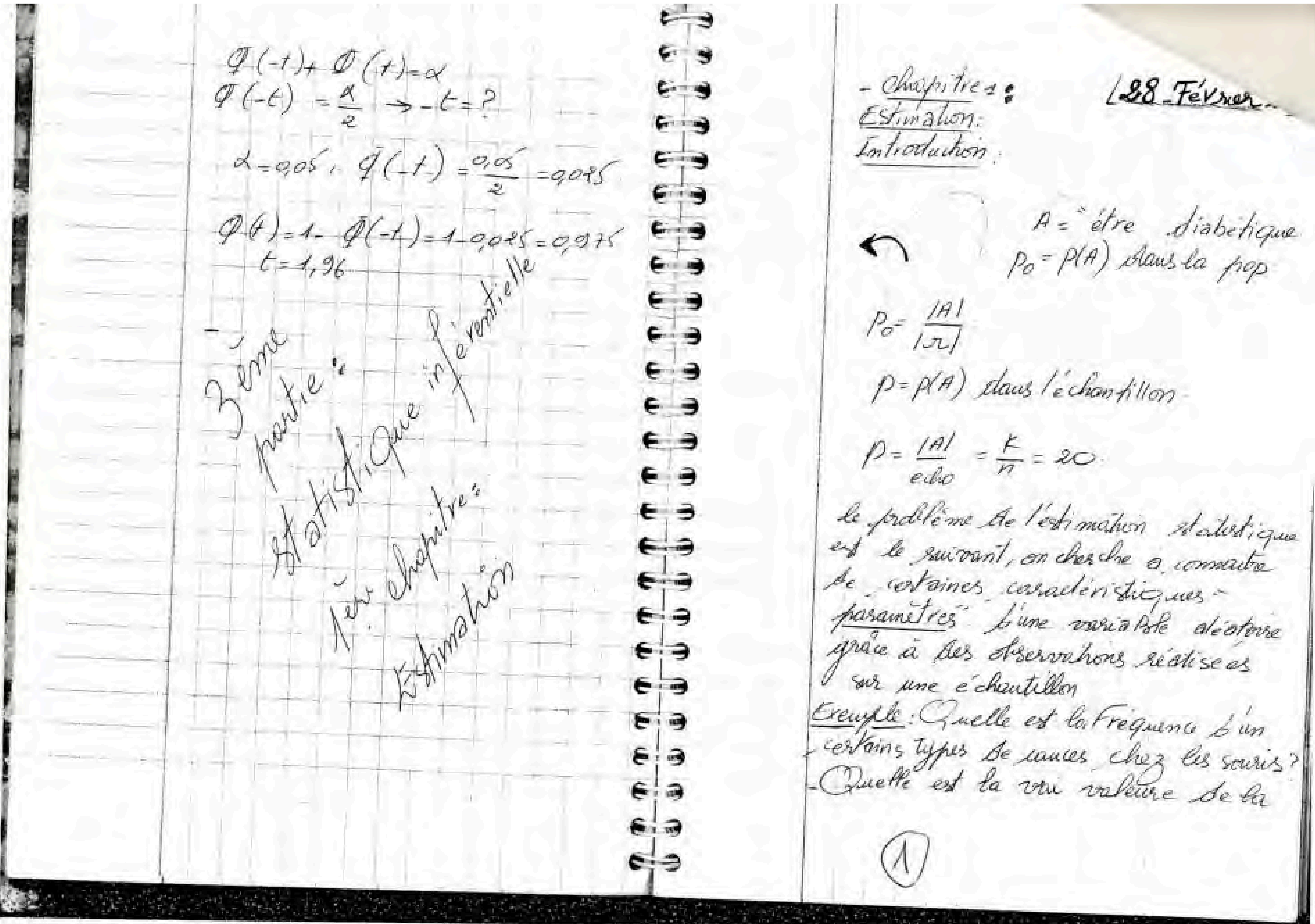
In the name of Allah, the Most Gracious, the Most Merciful









$\Phi(-t) + \Phi(t) = \alpha$

$\Phi(-t) = \frac{\alpha}{2} \rightarrow -t = ?$

$\alpha = 0{,}05 ; \Phi(-t) = \frac{0{,}05}{2} = 0{,}025$

$\Phi(t) = 1 - \Phi(-t) = 1 - 0{,}025 = 0{,}975$

$t = 1{,}96$

- 3ème partie :
Statistique inférentielle
1ier chapitre :
Estimation

- Chapitre 1 : 28 Février

Estimation :

Introduction :

A = " être diabétique "

$P_0 = P(A)$ dans la pop

$P_0 = \frac{|A|}{|\Omega|}$

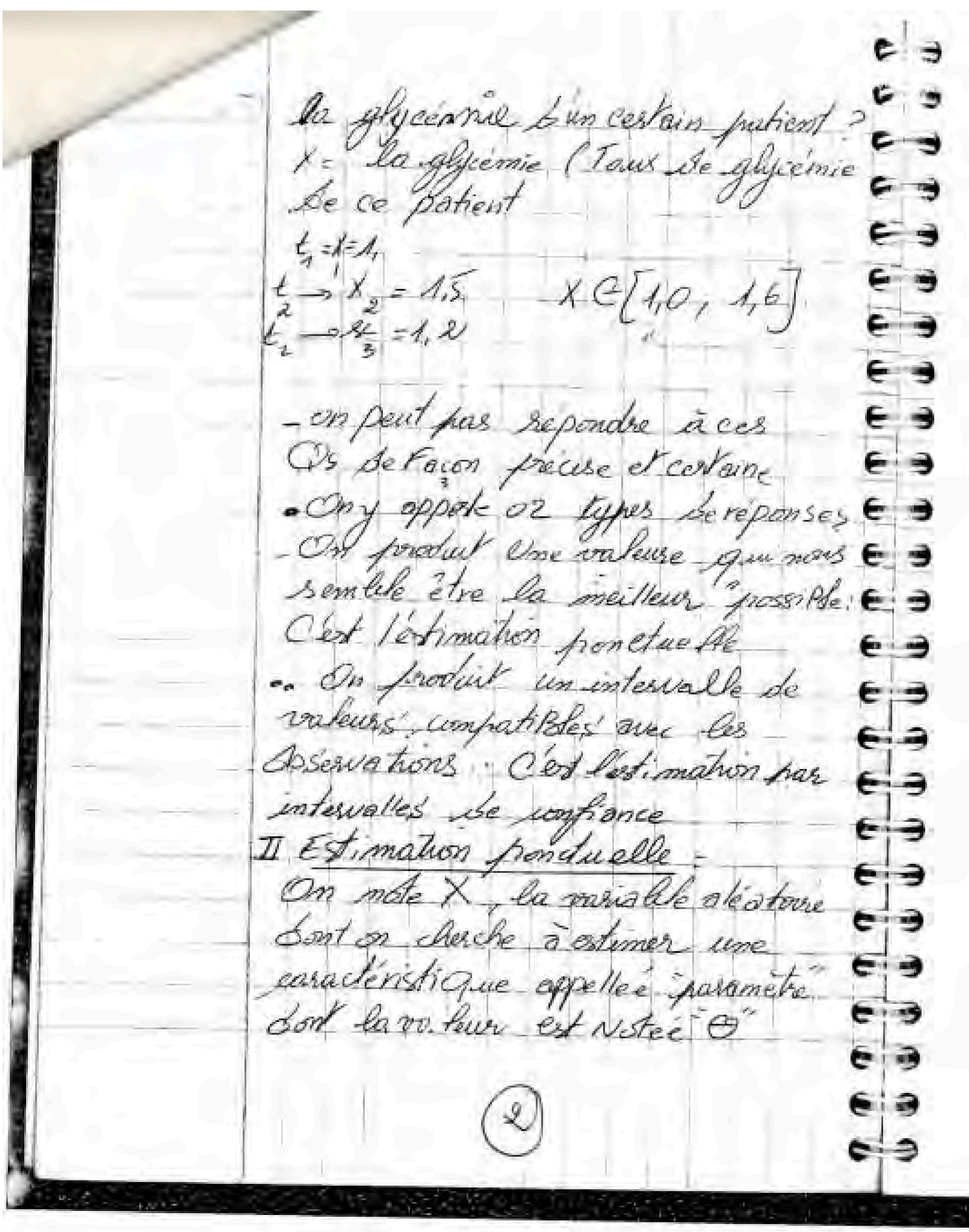
$p = p(A)$ dans l'échantillon

$p = \frac{|A|}{echo} = \frac{K}{n} = 20$

Le problème de l'estimation statistique est le suivant, on cherche a connaitre de certaines caractéristiques - paramètres " d'une variable aléatoire grâce à des observations réalisées sur une échantillon

Exemple : Quelle est la Frequence d'un certains types de cancer chez les souris ?
- Quelle est la vrai valeure de la

(1)



la glycémie d'un certain patient ?
X = la glycémie (Taux de glycémie de ce patient

$t_1 = X_1 = 1,1$
$t_2 \to X_2 = 1,5$ $X \in [1,0 ; 1,6]$
$t_3 \to X_3 = 1,2$

- on peut pas répondre à ces Qs de façon précise et certaine
- On y appelle 02 types de réponses
- On produit une valeur qui nous semble être la meilleur possible : C'est l'estimation ponctuelle
- On produit un intervalle de valeurs compatibles avec les observations : C'est l'estimation par intervalles de confiance

II Estimation ponctuelle :
On note X, la variable aléatoire dont on cherche à estimer une caractéristique appellée "paramètre" dont la valeur est notée "Θ"

(2)

X = Taux de Glycémie
$E(X) = m$ la glycémie moyenne
$m = \Theta$

1) Def d'un estimateur :
Soit X la v.a d'un interet et Θ est le paramètre à estimer
- Soit $(X_1, \dots, X_n)$ Un échantillon aléatoire représentatif extrait de la population
- Un estimateur T de paramètre Θ est fonction exclusive de l'échan (des observations)

Exemple : $E(X) = m$
$T = \bar{X}$ est un estimateur de m

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \dots x_n)$$

$f : \mathbb{R} \to \mathbb{R}$
$x \to f(x) = ax^2 + b$
(x_1, x_2, x_3)

propriété de sans biais

(3)

Soit T un estimateur de θ
le biais de T noté $B(T) = E(T - \theta)$
$\Rightarrow E(T) - E(\theta) \Rightarrow B(T) = E(T) - \theta$

Un estimateur T de θ est dit sans biais (E.S.B), ssi $\Rightarrow B(T) = 0 \Leftrightarrow E(T) = \theta$
Soient T_1 et T_2, 2 estimateurs de θ
Si $B(T_1) < B(T_2) \Rightarrow T_1$ est un meilleure que T_2 "

③ Exemples :
On note les paramètres
$m = E(X) \Rightarrow$ moy dans la pop
$\sigma^2 \Rightarrow V(X) \Rightarrow$ Var " " "
$p_0 = p(A) \Rightarrow$ la prévalence de A (pop) (probabilité de A = pop
On note par :
$\bar{x}$ = la moy de l'échantillon

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

S^2 = variance de l'échantillon

④

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 -$$

p = la probabilité de A dans l'écha

$$p = \frac{|A|}{n} \quad \frac{k}{n} = p$$

→ On estimera m par $\bar{x}$, on écrit
$\hat{m} = \bar{x}$
↳ la valeur estimée de m
$T = \bar{X}$ est l'estimateur de m v.a
→ On estimera σ^2 par S^2 mais pourtant
S^2 ∄ E.S.B de σ^2 · $E(S^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$

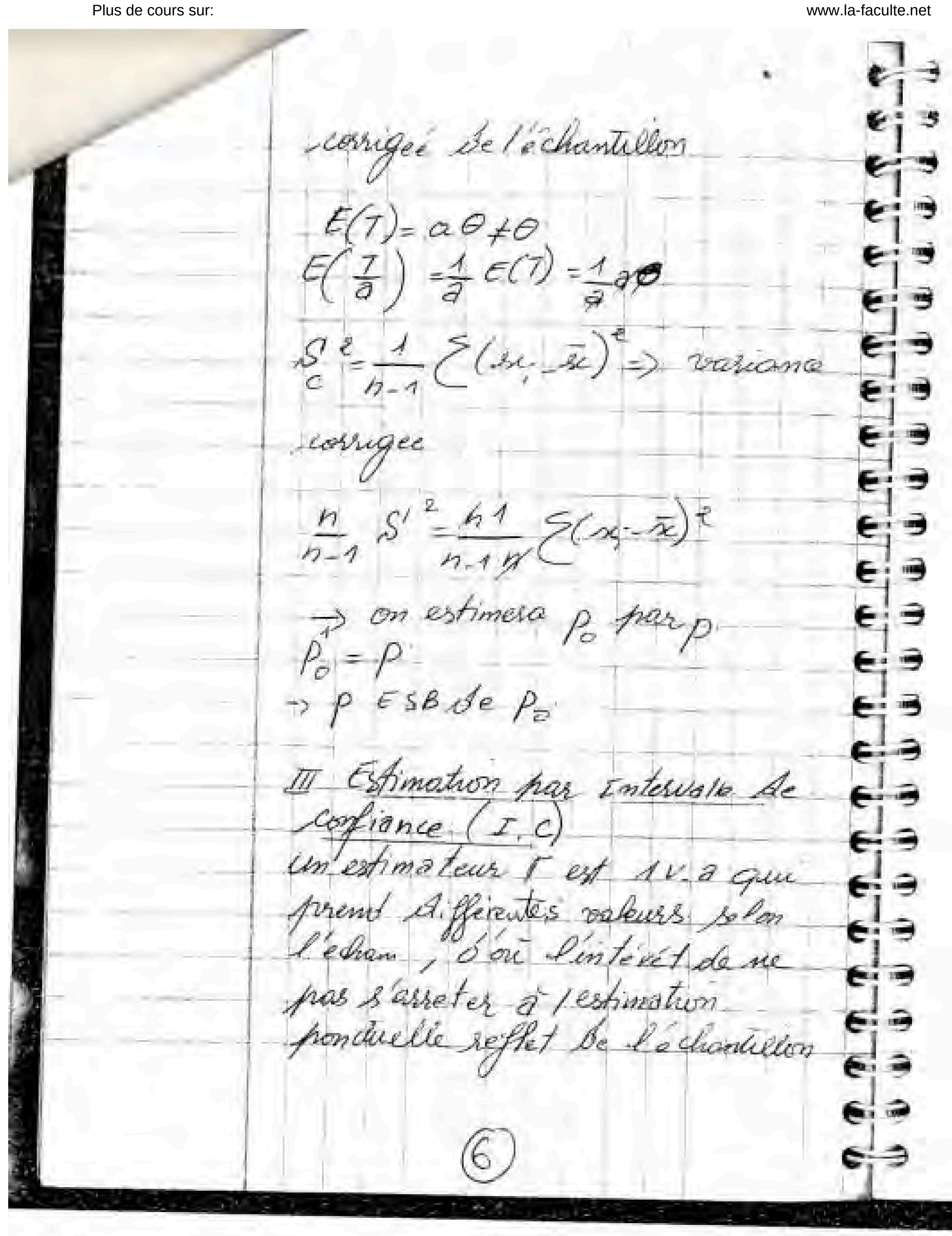
alors on preferera estimer σ^2 par $S_c^2 = \frac{n}{n-1}S^2$ qui est ESB de σ^2,

$$\sigma^2 = \frac{n}{n-1}S^2$$

$$\sigma^2 = S_c^2$$

$$S_c^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 \text{ : variance}$$

⑤



corrigée de l'échantillon

$E(T) = a\theta \neq \theta$

$E\left(\frac{T}{a}\right) = \frac{1}{a} E(T) = \frac{1}{a} a\theta$

$S_c^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \Rightarrow$ variance corrigée

$\frac{n}{n-1} S'^2 = \frac{n}{n-1} \frac{1}{n} \sum (x_i - \bar{x})^2$

→ on estimera p_0 par p

$\hat{p}_0 = p$

→ p ESB de p_0

III Estimation par intervalle de confiance (I.C)

Un estimateur T est 1 v.a qui prend différentes valeurs selon l'échan, d'où l'intérêt de ne pas s'arrêter à l'estimation ponctuelle reflet de l'échantillon

⑥

mais de trouver un intervalle $[a, b]$ qui a de forte chance de contenir la vrai valeur du paramètre θ

Déf. estimer un paramètre θ par IC $[a,b]$ au seuil α, $\alpha \in]0,1[\Leftrightarrow$

$P(\theta \in [a,b]) = P(a \leq \theta \leq b) = 1-\alpha$

$1-\alpha$ = niveau de confiance = Taux de sécurité (T.S)

α = risque

α : fixé au préalable, $\alpha = 0{,}05$

$\alpha = 0{,}01$, $\alpha = 0{,}02$ ... $\alpha = 0{,}1$

$\alpha = 1\%$, 2%, 10%.

② I de C pour la moyenne :

Trouver a et b tel que $P(a < m < b) = 1-\alpha$

a) supposons que σ^2 est connue : soit $T = f(x, \theta)$, T.q $T \rightsquigarrow$ Table

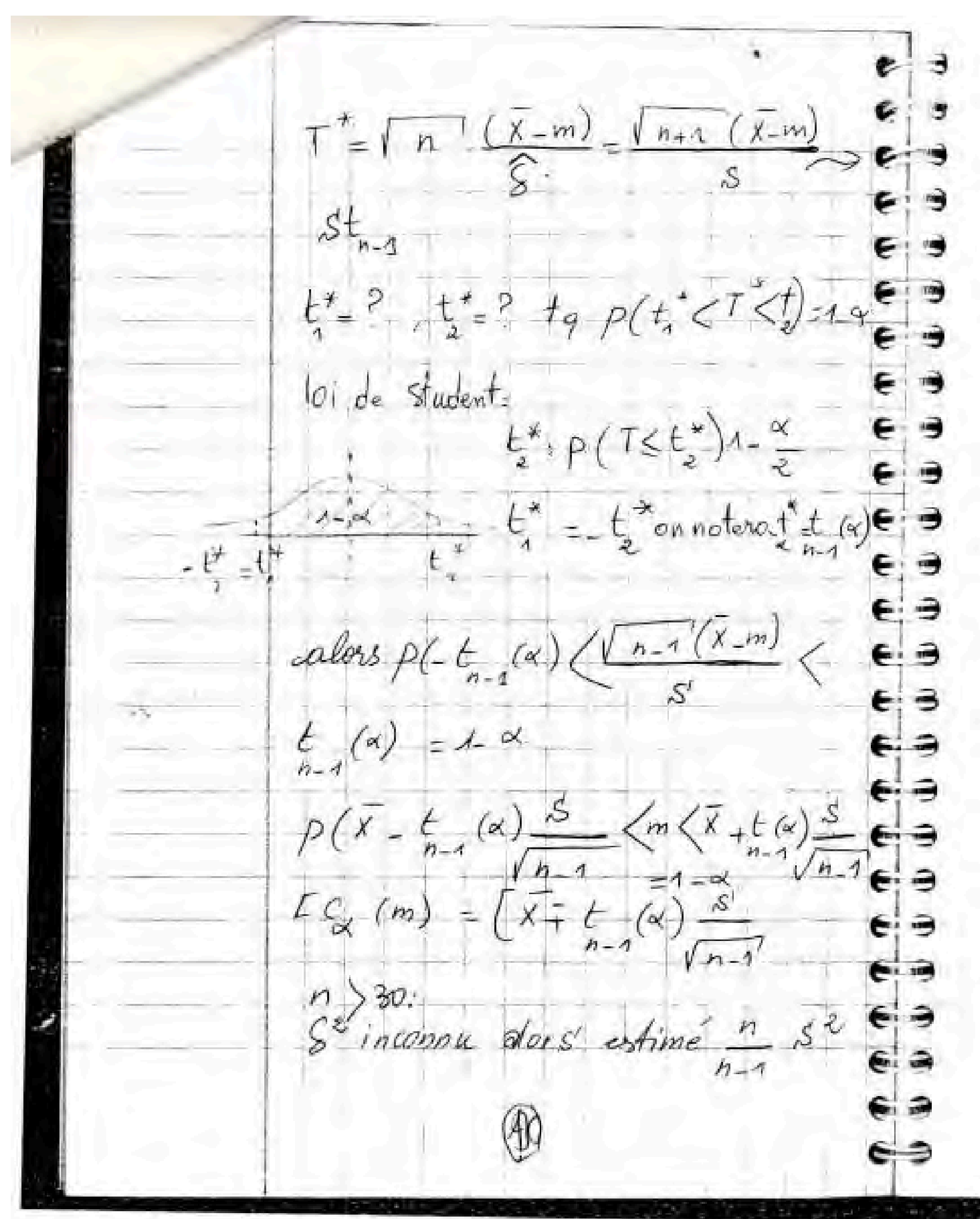
exp $N(0,1)$

$t_1 = ?$, $T_2 = ?$

$P(T_1 < T < T_2) = 1-\alpha$

exp : $T \rightsquigarrow N(0,1)$

⑦



$$T^* = \sqrt{n}\,\frac{(\bar X - m)}{\hat S} = \frac{\sqrt{n-1}\,(\bar X - m)}{S} \rightsquigarrow St_{n-1}$$

$t_1^* = ?$, $t_2^* = ?$ tq $P(t_1^* < T^* < t_2^*) = 1-\alpha$

loi de student:

t_2^* : $P(T \le t_2^*) = 1-\frac{\alpha}{2}$

$t_1^* = -t_2^*$ on notera $t_\alpha^* = t_{n-1}(\alpha)$

alors $P\left(-t_{n-1}(\alpha) < \frac{\sqrt{n-1}\,(\bar X - m)}{S'} < t_{n-1}(\alpha)\right) = 1-\alpha$

$$P\left(\bar X - t_{n-1}(\alpha)\frac{S}{\sqrt{n-1}} < m < \bar X + t_{n-1}(\alpha)\frac{S}{\sqrt{n-1}}\right) = 1-\alpha$$

$$IC_\alpha(m) = \left[\bar X \mp t_{n-1}(\alpha)\frac{S'}{\sqrt{n-1}}\right]$$

$n > 30$:

σ^2 inconnu alors s' estimé $\frac{n}{n-1}S^2$

(10)

$$\hat S^2 = \frac{n}{n-1}S'^2 \Rightarrow \hat S = \sqrt{\frac{n}{n-1}}\,S$$

$$\Rightarrow T^* = \frac{\sqrt{n}\,(\bar X - m)}{\hat S} = \frac{\sqrt{n-1}\,(\bar X - m)}{S}$$

$$\rightsquigarrow St_{n-1} = N(0,1)$$

$$P\left(q_{\frac{\alpha}{2}} < T^* < q_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(-q_{1-\frac{\alpha}{2}} < \frac{\sqrt{n-1}\,(\bar X - m)}{S} < q_{1-\frac{\alpha}{2}}\right)$$

$$\Rightarrow IC_\alpha(m) = \left[\bar X \pm q_{1-\frac{\alpha}{2}}\frac{S'}{\sqrt{n-1}}\right]$$

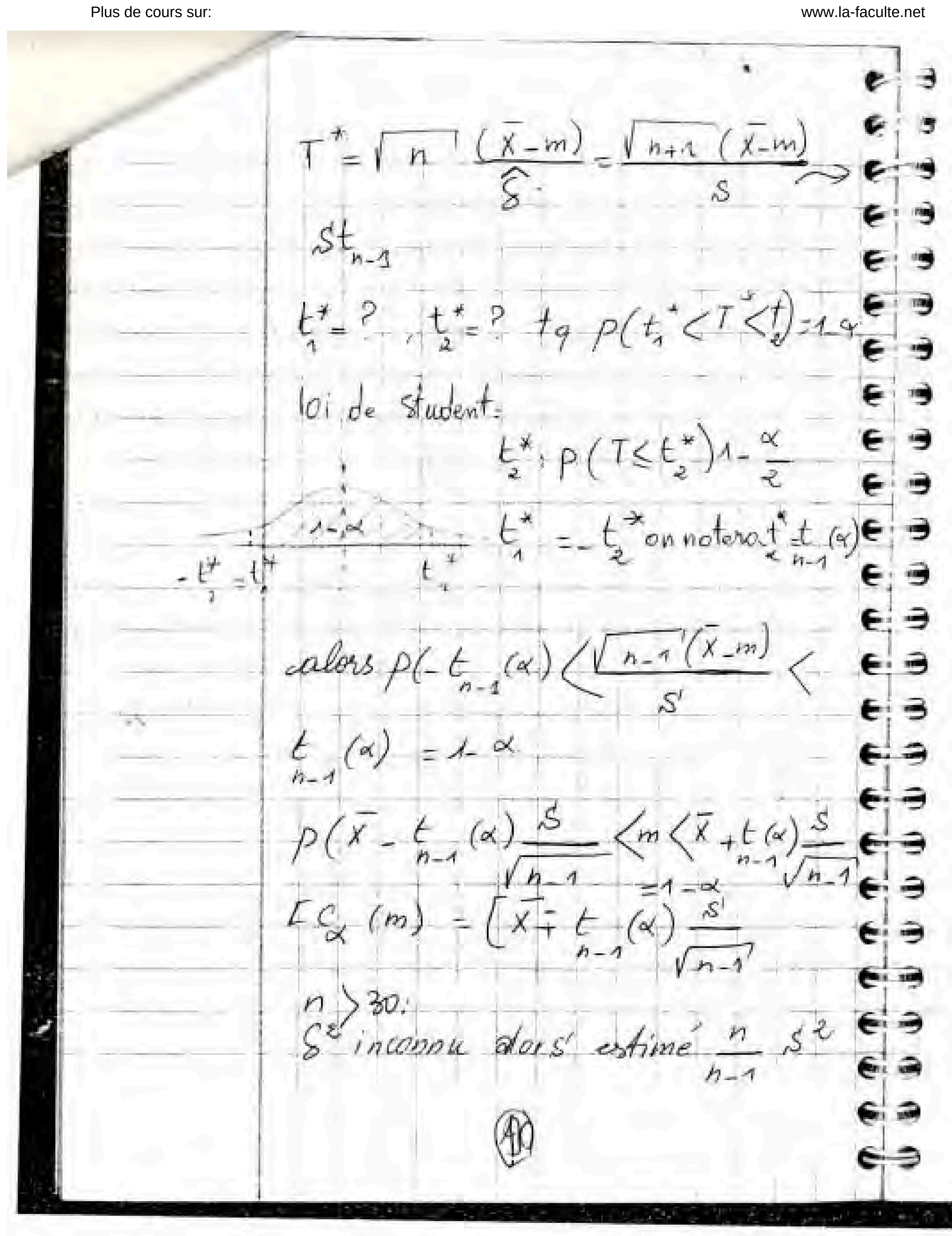
exp: on veut estimer le taux moyen de cholestérol d'un individu dans une population donnée

- Un échantillon de 16 personne a donné une moyenne de 1,8 et un écart type = 2,2 = S'
- donnez l'intervalle "IC" ?

$IC_\alpha(m) = ?$ $n \le 30$ et σ : inconnu

$\alpha = 0{,}05$

(11)



$T^* = \sqrt{n}\,\frac{(\bar{X}-m)}{\hat{S}} = \frac{\sqrt{n+1}\,(\bar{X}-m)}{S} \rightsquigarrow$

$\rightsquigarrow St_{n-1}$

$t_1^* = ?\ ,\ t_2^* = ?$ tq $P(t_1^* < T^* < t_2^*) = 1-\alpha$

loi de student:

t_2^* ; $P(T \leq t_2^*) = 1 - \frac{\alpha}{2}$

$t_1^* = -t_2^*$ on notera $t_\alpha^* = t_{n-1}(\alpha)$

alors $P\left(-t_{n-1}(\alpha) < \frac{\sqrt{n-1}\,(\bar{X}-m)}{S'} < t_{n-1}(\alpha)\right) = 1-\alpha$

$P\left(\bar{X} - t_{n-1}(\alpha)\frac{S}{\sqrt{n-1}} < m < \bar{X} + t_{n-1}(\alpha)\frac{S}{\sqrt{n-1}}\right) = 1-\alpha$

$IC_\alpha(m) = \left[\bar{X} \mp t_{n-1}(\alpha)\frac{S'}{\sqrt{n-1}}\right]$

$n > 30$:

S^2 inconnu alors s' estimé $\frac{n}{n-1} S^2$

(10)

$S^2 = \frac{n}{n-1} S'^2 \Rightarrow \hat{S} = \sqrt{\frac{n}{n-1}}\cdot S$

$\Rightarrow T^* = \frac{\sqrt{n}\,(\bar{X}-m)}{\hat{S}} = \frac{\sqrt{n-1}\,(\bar{X}-m)}{S'}$

$\rightsquigarrow St_{n-1} = N(0,1)$

$P\left(q_{\frac{\alpha}{2}} < T^* < q_{1-\frac{\alpha}{2}}\right) = 1-\alpha$

$P\left(-q_{1-\frac{\alpha}{2}} < \frac{\sqrt{n-1}\,(\bar{X}-m)}{S} < q_{1-\frac{\alpha}{2}}\right)$

$\Rightarrow IC_\alpha(m) = \left[\bar{X} \pm q_{1-\frac{\alpha}{2}} \frac{S'}{\sqrt{n-1}}\right]$

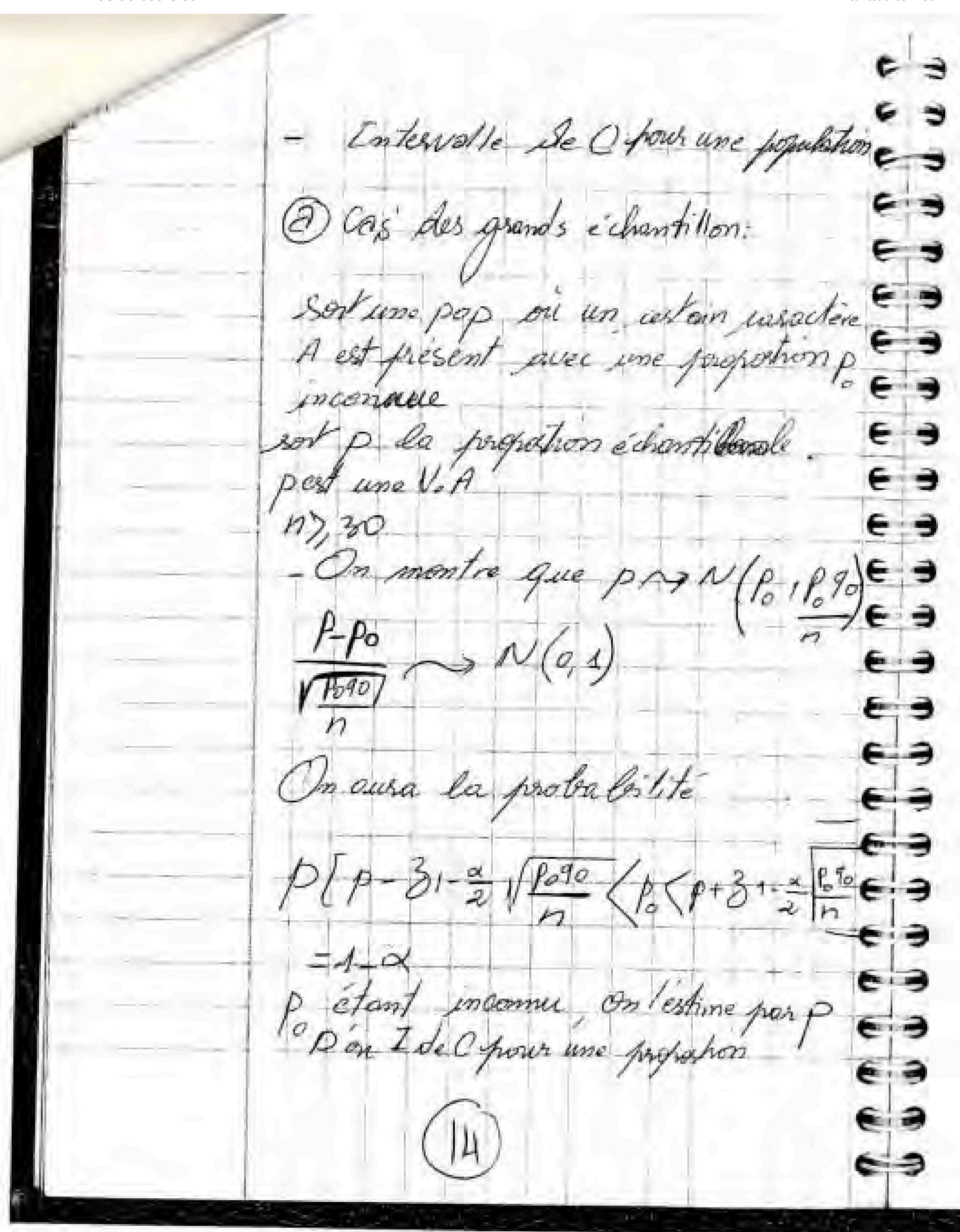
exp: on veut estimer le taux moyen de cholestérol d'un individu dans une population donnée

- Un échantillon de 16 personne a donné une moyenne de 1,8 et un écart-type = 2, 2 = S'
- donnez l'intervalle "IC" ?

$IC_\alpha(m) = ?$ $\quad n \leq 30$ et S : inconnu

$\alpha = 0{,}05$

(11)



- Intervalle de C pour une population

ⓐ Cas des grands échantillon:

Soit une pop où un certain caractère A est présent avec une proportion p_0 inconnue

soit p la proportion échantillonnale.

p est une V.A

$n \geq 30$

- On montre que $p \rightsquigarrow N\left(p_0, \frac{p_0 q_0}{n}\right)$

$$\frac{p-p_0}{\sqrt{\frac{p_0 q_0}{n}}} \rightsquigarrow N(0,1)$$

On aura la probabilité

$$P\left[p - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0 q_0}{n}} < p_0 < p + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0 q_0}{n}}\right] = 1-\alpha$$

p_0 étant inconnu, on l'estime par p

D'où I de C pour une proportion

(14)

$$I\ de\ C(p_0) = p \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}$$

b) cas des petits échantillons: $n<30$

- Il existe des tables (abaques) (Hors programme)

3) I de C pour une variance:

soit $X \rightsquigarrow N(m, \sigma^2)$, $I\ de\ C(\sigma^2) = ?$

soit (m connue)

On montre que $Q = \sum_{i=1}^{n}\left(\frac{x_i - }{\sigma}\right)^2 \rightsquigarrow \chi^2_1$ (loi du Khi Deux à 1

a, b ? tq $P(a<Q<b) = 1-\alpha$

$a = q_1$ est tel que $P(Q<q_1) = \frac{\alpha}{2}$

$q_1 = \chi^2_n\left(\frac{\alpha}{2}\right)$ [illegible]

$b = q_2$ est tel que $P(Q>q_2) = \frac{\alpha}{2}$

$q_2 = \chi^2_m\left(1-\frac{\alpha}{2}\right)$

(15)

$$P\left(q_1 < \frac{\sum (x_i - m)^2}{\sigma^2} < q_2\right) = 1-\alpha$$

$$= P\left(\frac{\sum (x_i - m)^2}{q_2} < \sigma^2 < \frac{\sum (x_i - m)^2}{q_1}\right) = 1-\alpha$$

I de C (σ^2) inconnu $= \left[\frac{\sum (x_i - m)^2}{q_2} ; \frac{\sum (x_i - m)^2}{q_1}\right]$

- Remarque : $n > 30$

$\chi^2_n \simeq N(n, 2n)$ n (nombre de d. L)

$Q \rightsquigarrow \chi^2_n$ q_2 / $P(Q < q_2) = 1-\alpha$

$$= P\left(\frac{Q-n}{\sqrt{2n}} < \frac{q_2 - n}{\sqrt{2n}}\right) =$$

$$\phi\left(\frac{q_2 - n}{\sqrt{2n}}\right) = 1-\alpha$$

(b) - m connu :

On montre que $Q = n\frac{S^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$

(16)

$$P(q_1 < Q < q_2) = 1-\alpha$$

$$P(Q < q_1) = P(Q > q_2) = \frac{\alpha}{2}$$

$$1-\alpha = P\left(q_1 < \frac{nS^2}{\sigma^2} < q_2\right)$$

$$= P\left(\frac{nS^2}{q_2} < \sigma^2 < \frac{nS^2}{q_1}\right) = 1-\alpha$$

